



## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/901,484A

TIME: 10:08:34

Input Set : D:\GEN-T111XC3D2.ST25.txt

Output Set: N:\CRF3\03052002\I901484A.raw

3 <110> APPLICANT: Cohen, Daniel  
 4 Blumenfeld, Marta  
 5 Chumakov, Ilya  
 6 Bougueleret, Lydie  
 8 <120> TITLE OF INVENTION: Prostate Cancer Gene  
 10 <130> FILE REFERENCE: GEN-T111XC3D2  
 12 <140> CURRENT APPLICATION NUMBER: US 09/901,484A  
 13 <141> CURRENT FILING DATE: 2001-07-09  
 15 <150> PRIOR APPLICATION NUMBER: US 08/996,306  
 16 <151> PRIOR FILING DATE: 1997-12-22  
 18 <150> PRIOR APPLICATION NUMBER: US 60/099,658  
 19 <151> PRIOR FILING DATE: 1998-09-09  
 21 <150> PRIOR APPLICATION NUMBER: US 09/218,207  
 22 <151> PRIOR FILING DATE: 1998-12-22  
 24 <150> PRIOR APPLICATION NUMBER: US 09/338,907  
 25 <151> PRIOR FILING DATE: 1999-06-23  
 27 <150> PRIOR APPLICATION NUMBER: US 09/853,526  
 28 <151> PRIOR FILING DATE: 2001-05-11  
 30 <160> NUMBER OF SEQ ID NOS: 578  
 32 <170> SOFTWARE: PatentIn version 3.1  
 34 <210> SEQ ID NO: 1  
 35 <211> LENGTH: 56516  
 36 <212> TYPE: DNA  
 37 <213> ORGANISM: Homo sapiens  
 39 <220> FEATURE:  
 40 <221> NAME/KEY: promoter  
 41 <222> LOCATION: (1629)..(1870)  
 42 <223> OTHER INFORMATION: identification method Proscan  
 43 <220> FEATURE:  
 44 <221> NAME/KEY: misc\_feature  
 45 <222> LOCATION: (1998)..(2000)  
 46 <223> OTHER INFORMATION: potential start codon  
 51 <220> FEATURE:  
 52 <221> NAME/KEY: misc\_feature  
 53 <222> LOCATION: (2001)..(2216)  
 54 <223> OTHER INFORMATION: exon 1  
 57 <220> FEATURE:  
 58 <221> NAME/KEY: misc\_feature  
 59 <222> LOCATION: (11994)..(14337)

ENTERED

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/901,484A

TIME: 10:08:34

Input Set : D:\GEN-T111XC3D2.ST25.txt

Output Set: N:\CRF3\03052002\I901484A.raw

66 <223> OTHER INFORMATION: Tyr Phos  
 69 <220> FEATURE:  
 70 <221> NAME/KEY: primer\_bind  
 71 <222> LOCATION: (11930)..(11947)  
 72 <223> OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID 42  
 75 <220> FEATURE:  
 76 <221> NAME/KEY: allele  
 77 <222> LOCATION: (12057)..(12103)  
 78 <223> OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID 24  
 81 <220> FEATURE:  
 82 <221> NAME/KEY: primer\_bind  
 83 <222> LOCATION: (12339)..(12358)  
 84 <223> OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID 51, complement  
 87 <220> FEATURE:  
 88 <221> NAME/KEY: primer\_bind  
 89 <222> LOCATION: (13547)..(13564)  
 90 <223> OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID 64  
 93 <220> FEATURE:  
 94 <221> NAME/KEY: allele  
 95 <222> LOCATION: (13657)..(13703)  
 96 <223> OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID 58  
 99 <220> FEATURE:  
 100 <221> NAME/KEY: primer\_bind  
 101 <222> LOCATION: (13962)..(13981)  
 102 <223> OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID 67, complement  
 105 <220> FEATURE:  
 106 <221> NAME/KEY: misc\_feature  
 107 <222> LOCATION: (17997)..(18003)  
 108 <223> OTHER INFORMATION: n = a, c, g, or t.  
 111 <220> FEATURE:  
 112 <221> NAME/KEY: misc\_feature  
 113 <222> LOCATION: (18196)..(18265)  
 114 <223> OTHER INFORMATION: exon 2  
 117 <220> FEATURE:  
 118 <221> NAME/KEY: misc\_feature  
 119 <222> LOCATION: (19357)..(19357)  
 120 <223> OTHER INFORMATION: n = a, c, g, or t.  
 123 <220> FEATURE:  
 124 <221> NAME/KEY: misc\_feature  
 125 <222> LOCATION: (23717)..(23832)  
 126 <223> OTHER INFORMATION: exon 3  
 129 <220> FEATURE:  
 130 <221> NAME/KEY: misc\_feature  
 131 <222> LOCATION: (25571)..(25660)  
 132 <223> OTHER INFORMATION: exon 4  
 135 <220> FEATURE:

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/901,484A

TIME: 10:08:34

Input Set : D:\GEN-T111XC3D2.ST25.txt

Output Set: N:\CRF3\03052002\I901484A.raw

141 <220> FEATURE:  
142 <221> NAME/KEY: allele  
143 <222> LOCATION: (34469)..(34515)  
144 <223> OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID 25  
147 <220> FEATURE:  
148 <221> NAME/KEY: primer\_bind  
149 <222> LOCATION: (34625)..(34645)  
150 <223> OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID 52, complement  
153 <220> FEATURE:  
154 <221> NAME/KEY: misc\_feature  
155 <222> LOCATION: (34669)..(34759)  
156 <223> OTHER INFORMATION: exon 5  
159 <220> FEATURE:  
160 <221> NAME/KEY: misc\_feature  
161 <222> LOCATION: (39083)..(39093)  
162 <223> OTHER INFORMATION: n = a, c, g, or t.  
165 <220> FEATURE:  
166 <221> NAME/KEY: misc\_feature  
167 <222> LOCATION: (40688)..(40846)  
168 <223> OTHER INFORMATION: exon 6  
171 <220> FEATURE:  
172 <221> NAME/KEY: misc\_feature  
173 <222> LOCATION: (48070)..(48193)  
174 <223> OTHER INFORMATION: exon 7  
177 <220> FEATURE:  
178 <221> NAME/KEY: misc\_feature  
179 <222> LOCATION: (50182)..(54523)  
180 <223> OTHER INFORMATION: exon 8  
183 <220> FEATURE:  
184 <221> NAME/KEY: primer\_bind  
185 <222> LOCATION: (51149)..(51168)  
186 <223> OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID 65  
189 <220> FEATURE:  
190 <221> NAME/KEY: allele  
191 <222> LOCATION: (51445)..(51494)  
192 <223> OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID 59  
195 <220> FEATURE:  
196 <221> NAME/KEY: primer\_bind  
197 <222> LOCATION: (51482)..(51499)  
198 <223> OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID 68, complement  
201 <220> FEATURE:  
202 <221> NAME/KEY: primer\_bind  
203 <222> LOCATION: (51596)..(51613)  
204 <223> OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID 44  
207 <220> FEATURE:  
208 <221> NAME/KEY: allele

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/901,484A

TIME: 10:08:34

Input Set : D:\GEN-T111XC3D2.ST25.txt

Output Set: N:\CRF3\03052002\I901484A.raw

```

214 <221> NAME/KEY: primer_bind
215 <222> LOCATION: (51996)..(52015)
216 <223> OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID 53, complement
219 <220> FEATURE:
220 <221> NAME/KEY: polyA_signal
221 <222> LOCATION: (54445)..(54450)
222 <223> OTHER INFORMATION: AATAAA
225 <400> SEQUENCE: 1
226 gtqgactctgt gactgttcgc aggaagagag gagcgggagc aggacagaca ataatgata      60
228 gtcaggagct gggtttggag ataaagaggg aaacaagagaa agttaagttc tgtgttttca      120
230 tggcaaacat tgcacaaaag tttaacaactt cgtgactaac agtaatctgg ggtgattcac      180
232 aacaaaattta cacataaaca catatttact gactttatac acagcaatcc taacgtgaac      240
234 acagaacctg ctttatcttt tcgcacactg ttctagtgtg gagatgtctg gtctcagtta      300
236 aaqaaaagcat aagggagcatt agttgrgcac actgtccaca cccgtgactt ttttcacca      360
238 gtactaaacc tagtgcttct tacagtacag ggcaatgaca gccacagaaa gagagaagct      420
240 ccttttactg tgtaatgctt cctgctggcc ttcaaatact tgttacttga gagatctcca      480
242 ttcaactggc ttgtcccca aaggtcatca totaccaatg atgttggtat ttgatgttaa      540
244 tcatgtataa agaaagtagc taccatcctg gccctgatta gaacttccca ctgaaatacc      600
246 gtccctgccta aaggtagcac aggtttccat tatggtgggt gtggggaggg ggccgggaata      660
248 tatatatata tatatatata tatatatatg gtaaagcatt cggcattctt ttaaagtaca      720
250 actatccttg aaaagggtta catattaaac catttttacc acagccaaag gggaggagaa      780
252 agatccaaaa gtccctgtgga tctgctttta catcaataaa acagttatcc acccttcgta      840
254 gcttttagtg aaggctacaa aagtatgctt tttatggatt acacatgtgc acccaactac      900
256 tttaattact acagaaaaaa acgaggctcc ttattaaaaa aaaatcagaa acaagtccaa      960
258 cagactctga ggaaatgaag caagagtga tctgaaaaag gtctaataaa cagtatggaa      1020
260 atatccttgt gggattgttc ttcagctatg cataaacatg taattatcat cattaactgt      1080
262 atggggaaaa acacggaccc taattctgaa acacccctgg agcagagagc gggcaggagg      1140
264 ggctgtctgc cactcagagc ggaggctgag gaggcggcgt ccccttgcaa aggaactggc      1200
266 gtgagcagat ggggacactc gagctgcccc gcgacctggg ccgagctgac taacaacctg      1260
268 gccaggtgct ctgcaagaat tagacctccg ataacgttaa caccaccttt ctcaactgct      1320
270 taattgtgtg catcccggtg cccaggggct tgtgagcagc aggtgcccgt tccaggcagc      1380
272 tccagcgacc cttaaacctg accgcgcgca cgtccggccc gagggagcag aacaagaggg      1440
274 acccggaacc tccctcgggc agcacccacc ttcacccagt tccgtcagtc gccaccacct      1500
276 cccctccccc gtccgcagcc ggcccagctg gggagcatgc gcagtggccg gagccgggtt      1560
278 gcccgcgcca cagcaggtag ctgtaactga actgtccggc caaacccaaa aatcaagaga      1620
280 cgtgttattg ccgcagaggt ggaactatgg caacggcgca ccaatcagaa ggccggttgt      1680
282 tgcggggag cccctgccc cggcaggggg atgtggcgat gggtaggggt catgggggtg      1740
284 gacpatccct gagcatoga tccgggaggg ccgpggggtc ccttgctttg ccjccgggag      1800
286 ccgggcagcc agcccgccac tcgctacccc ggcccggggc ggccggcgcc cccatgcggc      1860
288 tgggggggga ggctgggagc gggtaggggg ccggggggcc cggggccggg cggtgattgg      1920
290 ccgcctgctg gccgcgaact aggcgcggga ggcgggcggg gagcgcaggg ggagctcgct      1980
292 gccgcagagc tgagaagatg ctgctgtccc tgggtgtcca caactactcc atgcgctacc      2040
294 tctgtcccag cgtctgtgct ctgggcaagg cgcaccacca cgtgttgccc tggggggtct      2100
296 gggggtgtgt ctccgcttcc ctgcggccgc gcttctacca agcgtgggac gacgggtgt      2160
298 actggttcta ccagagcatg gtgctcttct tcttcagaaa ttacacgggg gtccaggtga      2220
300 gccgctccc gctccgggtt ctccgctccc acccgagctc ccaggagccc ggaactctcc      2280

```

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/901,484A

TIME: 10:08:34

Input Set : D:\GEN-T111XC3D2.ST25.txt

Output Set: N:\CRF3\03052002\I901484A.raw

308	ccctgcctg	ggtctgatgc	tgttagcaa	agtgggtgca	gatgcacgtt	ttaaataata	2520
310	gggcacgcgt	ttagcagttt	ctggcctttg	gtccaaagag	gtggtcctgt	tggaacagat	2580
312	gggagacgtc	tacactccga	agtgcgcttt	tacagtgaac	tcttgaaaca	gaagtacaat	2640
314	tgggtcttgt	gttctttccc	ctggacaagt	gaaagctggg	cgaagaaatg	aatacatttg	2700
316	ttaacccgtag	aagccctaact	agatacaatt	cttgccaact	ttaactgggc	ttgaatgtgt	2760
318	gggtgatctg	ttgtctgatt	actttctttc	tgttactgtt	tctctgtaga	gattggatto	2820
320	gtagattaaa	cttgagaaaac	aaaccataaa	agtggaaggc	cctctttaac	agtaggtatt	2880
322	tgaagtgtta	taaaaaaaaa	aaaggtgaat	ttttctttta	tttctcagtt	tgaaagaaca	2940
324	gctttattct	tggttattcc	taatgtccac	ctagtcctct	tttaactttc	ttggtagggg	3000
326	tagggtggca	tggggaaatg	ggacgggtac	attttgtctt	tttaactttt	ttttttcca	3060
328	ccacagcag	ctgtttttac	ctgtgtgtca	gtcaggtact	atatttagtt	tgcagttgca	3120
330	ctgctgatcg	acccttgatg	gcccagtttg	gaagttgttt	ggggggaagg	aactaggaga	3180
332	ggccaggggc	tccatttaaa	ccagtgtctg	taagtgtctc	cttggaaggga	aaaaaagata	3240
334	ctgttccagg	tcatggtttc	ctggtagttg	acgtttaaaa	tgggctcat	ttaaaaattt	3300
336	caataattca	ggctaatttt	ttccctttat	atggtaactc	ccccaagttt	gtctaaatgt	3360
338	atgattttta	tcatgattaa	gtttttactt	ccacatcatg	tgacaactgg	cctgggatgg	3420
340	gatataagct	cagaacacaa	agtcattcac	ctgttaaaaa	aataattcta	tctgtggcgg	3480
342	gttatgttat	ttttgttcaa	agaggacaca	atatgatgca	gaataaccca	ttgaaggatt	3540
344	ttttggtttg	gcaagttctt	atttttttaa	atggtctgta	aacctagcag	tgtttctgaa	3600
346	attgcatacc	ttacctgatg	ttcagagatc	cgattttaact	cttgatttcc	cagcaagtga	3660
348	ttttgaaaac	atttaattca	atcattcccc	ccaccgtctg	ttcaaatcaa	aggaagtggc	3720
350	atccagcact	aattttcatg	catttatgaa	aggatgcctg	aggaccctta	agtataatto	3780
352	aaaattttgt	ttaatgtgtg	ttccttgatg	aagttcttta	ggagtcttag	aacgaactga	3840
354	ttgcccactg	atcatcaaat	gcaagttatg	aacattttaat	aaaaatttaa	aaccaagagt	3900
356	ttcttgttcc	tgcattttta	tttttattgt	atggagggga	caaataatta	ttttctgttt	3960
358	agtaacagag	cagggtattt	tgaattttat	agggtctttt	tctgcagttc	gggtttcctg	4020
360	tgtacacaaa	gctacctttc	aatatttttt	attgtttctg	ttaagattaa	atcaatagag	4080
362	gaataaatag	ctatcttcaa	acataagacc	caaaggaaaa	agattttatg	tgatgttctg	4140
364	tcaccttatt	ttttacctgt	gaactttgtc	cattaaacttt	gtcactgaga	tgttttgatt	4200
366	aaaattttta	gcttgctttt	cttgttttgt	taggacaactc	ttttttctct	gaattgtttt	4260
368	tatcagcttt	cgtttgcaag	gctagtgatg	attctcttgt	tctgtataaa	gtattgttga	4320
370	ctcattttctg	aagggagttt	tagtaattta	agaggttata	agttttttaa	taaaaggttt	4380
372	attaattttat	atatattaaa	gaggcatttt	aaaataaaaat	ttttttttaa	tgacattttt	4440
374	acaccttttca	actctaggtt	taaaaaataa	gtggttcaca	gtagttcttg	cagaagaata	4500
376	ttttctttta	catagaattt	ttaagctgaa	gagaagtagt	agtaggtcca	tgagattttat	4560
378	gatctgtgct	tggcaggtaa	acctgcttcc	aacaaaattta	gttggtattt	tcttggtatto	4620
380	tgggtaaaata	cctttttctt	cccagtttcc	actactttat	tttccatagt	atctctgaga	4680
382	tagagaaaata	tttcagtcag	tgtgtctaaa	attgttctct	ataactcgtt	tatcttttta	4740
384	ggctccttcca	gaatctctca	ttgttactga	aaotcaaatg	ggtaactttct	tcacatttta	4800
386	ttctttttaga	ataagtaata	agaattttat	aagctttttt	atatttcaag	taatttgaga	4860
388	ctatttgaaaa	tccagtttaag	tctctctact	gtgttgagag	gcatttgatto	aagtaacctg	4920
390	gttaactttcc	tgtgtctgca	aaacagatca	cctcaaaacta	agcgggttaa	aataatagaa	4980
392	cttaagttct	cgtgattctg	gaggccagca	ctttgaaatc	aaggtgtagg	ctcaatttta	5040
394	ctccctctgg	aggccctagg	gggaactctgt	tcttgtgggt	ttcaactttct	ggtgaactgt	5100
396	ggcatttctt	ggcttggggc	cccatcactt	caacctctgc	cttacagttc	ttgtgtccac	5160
398	ctcttctgtc	tcacatctca	ctctcccttt	ctcttagaag	gatgcttctc	attagattta	5220

## VERIFICATION SUMMARY

DATE: 03/05/2002

PATENT APPLICATION: US/09/901,484A

TIME: 10:08:35

Input Set : D:\GEN-T111XC3D2.ST25.txt

Output Set: N:\CRF3\03052002\I901484A.raw

L:824 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:826 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:870 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:1528 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:2372 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:3050 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:22  
L:3353 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:31  
L:4381 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:69  
L:4637 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:72  
L:4845 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:4847 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:4887 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:4993 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5021 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5023 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5025 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5029 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5087 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5107 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5109 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5137 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5173 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5243 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5275 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:5301 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:13038 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:182  
L:13086 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:182  
L:13297 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13327 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13369 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13371 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13485 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13565 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13579 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13685 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13861 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13863 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13867 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:13871 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183  
L:14541 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:183